

# Structure and Function of HIV-1 Integrase

Thang K. Chiu and David R. Davies\*

NIH, NIDDK, Laboratory of Molecular Biology, Building 5, Room 338, Bethesda, MD 20892-0560, USA



**Abstract:** HIV-1 integrase is a multidomain enzyme which is required for the integration of viral DNA into the host genome. It is one of three enzymes of HIV, the others being the Reverse Transcriptase and the Protease. It is an attractive target for therapeutic drug design. The enzyme consists of three domains. The N-terminal domain has a His<sub>2</sub>Cys<sub>2</sub> motif which chelates zinc, the core domain has the catalytic DDE motif which is required for its enzymatic activity, and the C-terminal domain has an SH3-like fold which binds DNA nonspecifically. We review the structures of various integrase fragments, the core domain with inhibitors bound, and propose a model for DNA binding.

## INTRODUCTION

HIV-1 integrase catalyzes the integration of the reverse transcribed viral DNA into the host genome. It is an essential enzyme for the production of new virus. As one of the three enzymes of HIV-1 it is an attractive target for chemical intervention.

The integration reaction is carried out in two steps: 3'-end processing and 3'-end joining or strand transfer (Fig. 1a), which integrase alone can carry out. In the processing reaction integrase removes the terminal GT dinucleotide from each of the viral DNA ends, leaving a recessed CA with a free 3' hydroxyl group and an overhanging 5'-AC on the complementary strand. In the strand transfer reaction integrase cleaves the target DNA, ligating both 3'-ends of the viral DNA to the 5' ends of the cleaved target DNA. Integrase also carries out a third reaction, termed the disintegration reaction, which is the reverse of the strand transfer reaction in which a substrate that mimics one end of the viral DNA joined to the target DNA is cleaved into its viral and target DNA parts [1-4].

Protease digestion and functional complementation studies show that HIV-1 integrase contains three domains [5-8] (Fig. 1b). The N-terminal domain (1-50) binds zinc, the core domain (50-212) contains the catalytic triad motif (D, D, 35E) characteristic of many polynucleotidyl transferases, and the C-terminal domain (213-288) binds DNA nonspecifically. The structures of each of these separate domains have been determined by X-ray diffraction [9-18] or by solution NMR [19-23]. Structures also exist for the core domain plus N-terminal domain [24] and for the core domain plus the C-terminal domain [25-27]. However, there are at present neither structures for the complete integrase molecule nor for integrase complexed with DNA substrate. There have also been structures determined for a number of other retroviral integrase fragments.

In this review we shall describe the integrase structures and discuss the various hypotheses relating these structures to the mechanism of action. For reasons of space we will focus specifically on the HIV-1 enzyme.

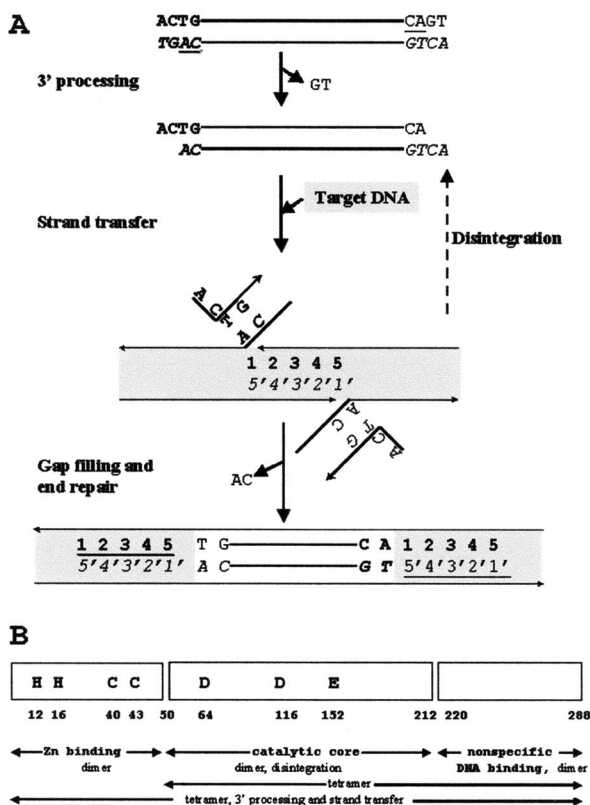
## 1. STRUCTURE OF THE CATALYTIC CORE DOMAIN

The core domain, residues 50-212, contains the catalytic triad motif D,D, 35E commonly found in polynucleotidyl transferases. Substitution of any of the three catalytic residues (D64, D116 and E152) in HIV-1 abolishes all three reactions [6,28-30]. The core domain alone can carry out the disintegration reaction [6,28,31-34], but both terminal domains are required for 3' processing and strand transfer [30,31,33,35-38]. X-ray diffraction analysis of the core domain shows it to consist of an  $\alpha/\beta$  structure containing a five strand  $\beta$ -sheet together with six  $\alpha$ -helices (Fig. 2a). The monomer is structurally similar to RNaseH, MuA transposase and RuvC resolvase [39-41]. Two core domains associate to form a two-fold axis related dimer with a large solvent excluded interface of about 1400 Å<sup>2</sup> per subunit. The crystallization of the core domain was initially impeded by its poor solubility. A systematic replacement of the hydrophobic residues resulted in a mutant, F185K, that had considerably improved solubility and had as much activity for the disintegration reaction as the wild type domain [42]. The full length protein with the F185K mutation carried out all three of the catalytic actions of the wild type enzyme *in vitro*, showing no significant differences. However this mutation was deleterious to the virion assembly *in vivo* [43,44]; in contrast, the F185H mutation has full activity and does not disrupt virion assembly *in vivo* [44]. With the exception of four F185H core structures [12,15] this F185K mutation has been preserved in all the HIV-1 integrase crystal structures.

The three catalytic residues of the core domain define the position of the active site (Fig. 2a). Several of the early structures, crystallized using cacodylate as buffer, showed

\*Address correspondence to this author at NIH, NIDDK, Laboratory of Molecular Biology, Building 5, Room 338, Bethesda, MD 20892-0560, USA; E-mail: david.davies@nih.gov

some distortions in the side chain positions in the vicinity of Asp116 due to the reactivity of cacodylate with Cys65. However, it has been demonstrated that the core domain in the presence of cacodylate and reducing agent still retains disintegration activity [45]. Crystallization in the absence of cacodylate produced new crystal forms which could be shown to bind one atom of  $Mg^{2+}$  [10,12].



**Fig. (1).** (A) Catalytic steps involved in the insertion of viral DNA into the human genome. Viral DNA is unshaded while target DNA is shaded in light grey. The left viral LTR is in bold while the right viral LTR is not. The conserved CA dinucleotide is underlined. Arrows in the double disintegration substrate show the 5' to 3' direction of each DNA strand. Integration of viral DNA into the host genome results in the duplication of a few bases of the host genome. Underlined bases in the integrated product indicate bases which are filled in by host repair enzymes. (B) Domain boundaries of HIV-1 integrase as defined by protease digestion, functional complementation, and crystal structure analysis.

A number of structures now exist for the HIV-1 core domain (Table 1). In overall topology these structures are all similar to the structure in Fig. (2a), yet there are differences in detail. In particular there is a loop, residues 140-152 that is adjacent to the active site region and is quite disordered in many of the structures and shows different conformations in those structures in which it is ordered. The most likely explanation for these differences is that this loop is flexible and only adopts a relevant conformation in the presence of DNA substrate. It has been shown that the mutations G140A and G149A make this loop more rigid but reduce the catalytic activity without impairing DNA binding [11].

## Comparison with other Core Domains

Crystal structures have been determined for several other retroviral integrase core domains which have overall structures very similar to HIV-1 integrase. The ASV core domain can carry out both the processing and disintegration reactions; however, the processing reaction of this isolated domain produces a trinucleotide instead of a dinucleotide [37]. The core domain of ASV has been extensively studied [13,14,16,17,47]. The structure is generally very similar to HIV-1 integrase core, except in those regions that are disordered in either of the two structures. The mode of dimerization is similar although not identical. The solvent excluded surface area in the ASV dimer interface is  $766 \text{ \AA}^2$  per monomer vs  $1395 \text{ \AA}^2$  for the HIV-1 core, consistent with its lower association constant [13]. Comparison of the active site residues shows a difference in the side chain orientation of Asp64. As observed with HIV-1 [11], rigidification of the flexible loop (residues ~140 – 150), but at lower pH instead of with G140A/G149A mutation, correlates with lower activity [47].

For the SIV core domain [26] the overall structure again resembles that of HIV-1 [9] and ASV [13]. In the four different structures in the asymmetric unit the Glu152 adopts several side chain conformations: it can point to and away from the Asp's depending on the conformation of the flexible loop. This loop is ill defined although Glu152 can be clearly seen in all four cases. The position of the C of Asp64 is the same in most but the side chain varies.

The crystal structure of an active two-domain derivative of RSV integrase has been reported [27]. The core domain and dimer interface are similar to the previously reported HIV-1, ASV and SIV domains.

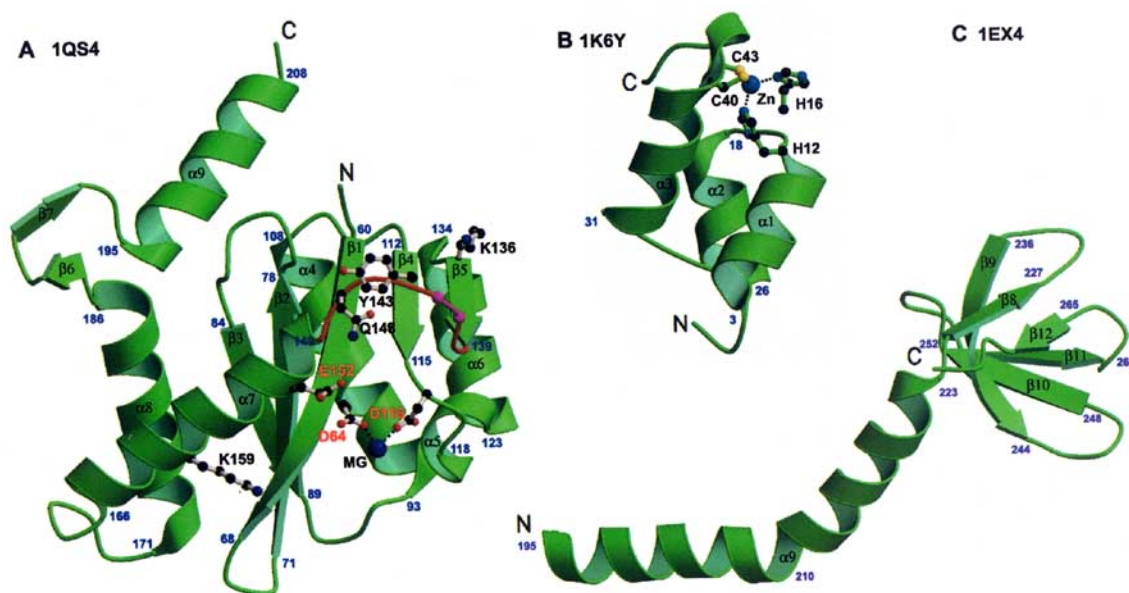
In conclusion, these four different retroviral core domains are all quite similar in structure, implying that these integrases almost certainly share the same mechanism of action. Differences in the locations of the side chains of the active site residues and of the conformations of the flexible loop are due at least partly to crystal packing, but most probably to the absence of DNA substrate in the crystal structures.

## 2. STRUCTURE OF THE N- AND C-TERMINAL DOMAINS

### The C-Terminal Domain

The C-terminal domain (213 – 288) is the least conserved of the three domains, binds DNA nonspecifically, and with the exception of FIV [34], deletion of this domain abolishes 3' processing and strand transfer activities [31,33-35,38,48-50]. Two NMR studies on solutions of domains consisting of residues 219-270 showed five strands arranged antiparallel to form a beta barrel which adopts an SH3-like fold [19,20]. In both cases the domains dimerize to form a symmetrical dimer. The two structures determined by NMR agree closely with each other.

The crystal structure of a multiple mutant of the two domain HIV-1 integrase, residues 52-288 has also been determined [25]. This structure, together with the N-terminal



**Fig. (2).** Individual domains of HIV-1 integrase are shown here, with secondary structures numbered sequentially from the N-terminus. Blue numbers give the amino acid position of the beginning and end of each secondary structure. The PDB accession number of the coordinate file is given above each structure.  $Mg^{2+}$  and  $Zn^{2+}$  cations are shown in blue. N and C denotes the beginning and end of each peptide fragment. (A) Catalytic core [46]. The flexible loop from residues 139 to 149 is shown in red, with magenta showing residues which are missing in the model. Catalytic residues and residues which have been shown to be crosslinked to integrase are depicted in ball-and-stick and are labelled in red and black characters, respectively. (B) N-terminus [24]. This His<sub>2</sub>Cys<sub>2</sub> domain binds  $Zn^{2+}$  and is involved in protein multimerization. (C) C-terminus [25]. The C-terminus from the core plus C-terminal domains structure is shown here as well as the leading  $\alpha$ -helix which joins it to the core domain. This domain has an SH3-like fold and consists of two antiparallel beta sheets and binds DNA nonspecifically.

domain from the core + N-terminus structure [24] superimposed, is shown in Fig. (3a). The core domain is similar to the other structures of the isolated core domain and the two-fold symmetrical dimer interface is the same. Two C-terminal domains are related to each other by 90° rotation relative to their two-fold axis and are connected to the core by helix 9, which are asymmetric for each monomer. A kink in the helix exists at the proteolytically-sensitive residue T210 on one of the monomer (Fig. 2a). The structure of the C-terminal domain monomer is similar to the structures observed by NMR. The dimer interface however is quite different in each of the three dimers observed in this crystal. One, the most similar to the NMR dimer, has the interface strands 9- 10- 11 running in opposite directions and has a CHAPS molecule in the interface contacting residues L242 and W243. Strands 9- 10- 11 also make hydrogen bonds and hydrophobic interactions with 9 of a two-fold related molecule. The other two dimer interfaces observed in this structure are also different from the NMR solution dimer. These results imply that the interaction between monomers can vary considerably and it is not at all clear what the preferred mode of dimerization would be for the C-terminal domain in the intact molecule.

For the RSV and SIV structures of the core plus C-terminus shown in Fig. (3), the C-terminus monomer is also similar to the HIV-1 monomer described above. In the SIV crystals only one C-terminal domain is observed among the four in the asymmetric unit and this interacts with a core

domain to form a possible surface for DNA binding [26]. In RSV the monomer is again very similar to that observed by NMR [27]. However in both crystal forms of the RSV 49-286 the C-terminal domains dimerize asymmetrically and their orientation relative to the core domain positions one of the domains much closer to the core domain.

The segment connecting core and C-terminal domains is different among HIV-1, SIV and RSV integrase: it's a  $\alpha$ -loop- in HIV-1, and loop- loop- in SIV and RSV. The  $\alpha$ -helix in HIV-1 (7 turns) is much longer than that of RSV (1 turn) and SIV (3 turns) and is kinked in the middle in one of the two monomer in both of the noncrystallographically related dimers. One of the C-terminal domains of RSV integrase does not contact its own core whereas the other one does, possibly an artifact of crystal packing [24].

### The N-Terminal Domain

The N-terminal domain (residues 1-50) of all retroviral integrases contains a His<sub>2</sub>Cys<sub>2</sub> motif, is highly conserved among integrases of all retroviruses and eukaryotic retrotransposons, and binds one equivalent of  $Zn^{2+}$  [51]. Although the N-terminal domain is required for 3' processing and strand transfer, its exact role in these reactions is not understood. Deletion of this domain in RSV [48] and Visna virus [8] integrases has no effect on 3' processing. Deletion in RSV integrase also has no effect on strand transfer, however, the virus can not replicate [48,52]. Substitution of His12 and His16 in RSV integrase does not significantly

**Table 1. Published Crystal Structures of Various Fragments of HIV-1 Integrase.**

NDB	Ref.	Res. (Å)	Space Group	Buffer & cation	Loop residues missing
<b>50-212, F185K</b>					
1itg	[9]	2.30	P <sub>3</sub> <sub>1</sub> 21	cadodylate	141-153
1b9d	[11]	1.70	P <sub>3</sub> <sub>1</sub> 21	cadodylate	141-148
1b92*	[11]	2.02	P <sub>3</sub> <sub>1</sub> 21	cadodylate	141-148
1b9f*	[11]	1.70	P <sub>3</sub> <sub>1</sub> 21	cadodylate	present
1biz	[10]	1.95	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub>	Hepes	140-148
1bis	[10]	1.95	P1	Hepes	A:143-150; B: present
1biu	[10]	2.50	C2	Hepes, Mg	A:141-147; B-C:141-148
1qs4	[46]	2.10	C2	Hepes, Mg	A:141-144; B:141-142; C:141-143
<b>50-212, F185H</b>					
1bi4	[12]	2.46	C2	Hepes	A:141-150; B:140-149; C: present
1bl3	[12]	2.00	C2	Hepes, Mg	A:141-150; B:140-149; C: present
1bhl	[12]	2.20	P <sub>3</sub> <sub>1</sub> 21	cadodylate	138-153
2itg	[15]	2.60	P <sub>3</sub> <sub>1</sub> 21	cadodylate	all present
<b>52-210, C56S, W131D, F139D, F185K, C280S</b>					
1exq	[25]	1.60	P <sub>3</sub> <sub>2</sub>	citrate, Cd	144-150, 142-150
1-212 W131D, F139D, F185K					
1k6y	[24]	2.40	P <sub>4</sub> <sub>3</sub> 2 <sub>1</sub> 2	acetate	A-C:140-148; D:140-152
<b>52-288 C56S, W131D, F139D, F185K, C280S</b>					
1ex4	[25]	2.80	P <sub>3</sub> <sub>1</sub> 2	citrate	142-144; 138-149

pH of acetate, citrate, cadodylate, and hepes buffer are 4.6, 5.6, 6.5, and 7.0 (1bis to 1qs4) or 7.5 (1bi4 and 1bl3), respectively. 1qs4 has 5CITEP inhibitor. \* 1b92 and 1b9f also contain G149A and (G140A + G149A) mutations, respectively.

impair 3' processing or strand transfer [53]. This is in contrast to HIV-1 integrase, where both histidines are required for both activities [6]. Mutation of the two cysteines in the His<sub>2</sub>Cys<sub>2</sub> motif in HIV-1 integrase also affects 3' processing and strand transfer [6]. With the exception of human foamy virus, integrase lacking this domain can perform disintegration [49]. Zinc promotes multimerization and enhances the catalytic activity of HIV-1 integrase [51,54]. Although this domain is involved in protein tetramerization in full length integrase, mutants lacking this domain can still form tetramers. Each domain alone exists in monomer – dimer equilibrium in solution [42,55] while the full-length or C-terminal plus catalytic core exists in dimer – tetramer equilibrium [43].

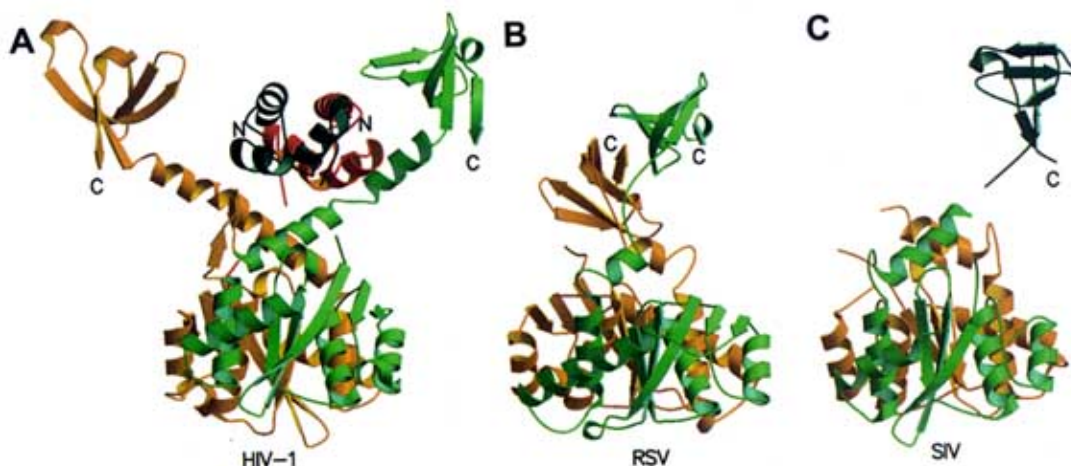
The NMR structure of the isolated N-terminal domain of HIV-1 and HIV-2 integrase in solution has been determined by two groups [21,23]. The structure is highly  $\alpha$ -helical with the monomer consisting of four helices (Fig 4b). The upper

region of the structure is stabilized by a hydrophobic core while the lower region is stabilized by the Zn<sup>2+</sup> coordination. There is an equilibrium between two conformations that differ in their arrangement around the Zn<sup>2+</sup>. The domain exists as a dimer with a subunit interface that excludes a solvent accessible area of ~550 Å<sup>2</sup> per subunit.

#### HIV-1 IN (1-212)

The structure of the N-terminal plus core domains (residues 1-212) for an HIV-1 triple mutant (W131D, F139D, F185K) has been determined [24]. The crystals contain four monomers per asymmetric unit. The core domain forms the same dimer as observed previously. The linker region joining the N-terminal and core domains (residues 47-55) is disordered in all four structures. However there appears to be no ambiguity in the assignment of the two N-terminal domains to the appropriate core domains. A pair of N-terminal domains approximately share a pseudo





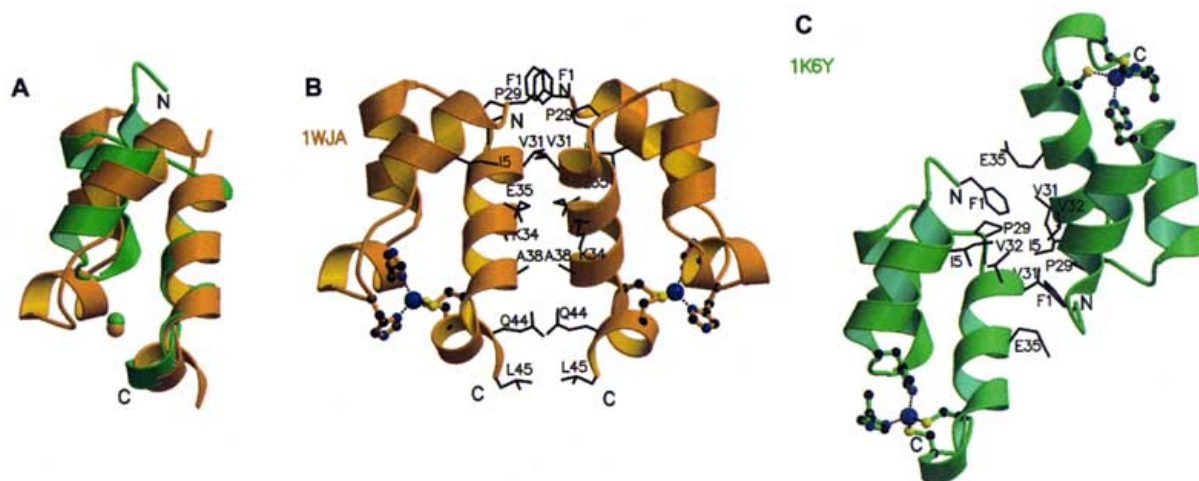
**Fig. (3).** The two-domain structures of HIV-1, RSV and SIV are shown here with each monomer in its own color. (A) HIV-1 (1-212) [24] and HIV-1 (50-288) [25] structures are superimposed to give a possible model of full length integrase. This resulting model surprisingly has no severe steric overlap of protein residues. (B) RSV (52-272) [27]. (C) SIV (50-293) [26]. The C-terminal domain is shown in darker green because it could geometrically be linked to either the green or the gold core domains.

two-fold axis with a pair of core domains to form two nearly identical dimers within the asymmetric unit (Fig. 3a and 4c). The two-fold axis in each dimer is only approximate and the angle between the N-terminal and core domains for the two subunits differs by 15 degrees.

The monomer structure of the N-terminal domain is very similar to the NMR structure of the isolated domain (Fig. 4a). However, the dimer interface is again quite different. In the NMR structure it is dominated by interactions between the third helix (Fig. 4b), but in the crystal structure it comprises the first and third helices, with a much smaller dimer interface (Fig. 4c). The N-terminal and core domains

are in contact, sharing a rather hydrophilic interface of total area 900 Å<sup>2</sup>. However, this contact occurs between domains contributed from separate dimers within the asymmetric unit, specifically residues 13 – 26 of the N-terminal domain of one dimer and residues 150 – 196 of the core domain of another dimer. The orientation of the N-terminal domain relative to the core is different by 15° between monomers A and B, but this difference could be real, as it is conserved among AB and CD dimers.

The four monomers in the asymmetric unit, AB and CD are in contact and are related by a noncrystallographic two-fold axis. The AB and CD interface is considerable and

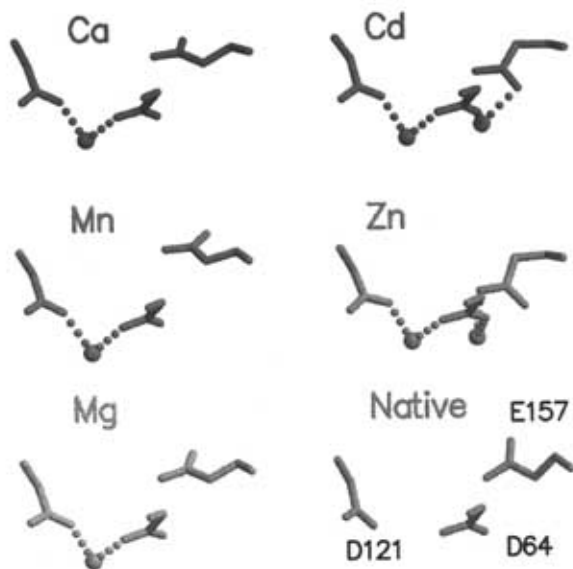


**Fig. (4).** Structures of the N-terminal domain determined by X-ray (green, 1K6Y) [24] and NMR (gold, 1WJA) [21]. (A) Residues 19 – 43 of the two structures can be superimposed, with RMSD values of 0.9 Å for backbone atoms. The main difference between the two models occurs in residues 9 – 19, which in the NMR structure results in the insertion of an  $\alpha$ -helix between the first and second  $\alpha$ -helices of the x-ray model. (B and C) The dimerization interface in the NMR and x-ray structures is completely different and involves residues F1, I5, P29, V31 and E35 in both structures and additionally residues V32 in the x-ray and K34, Q44 and L45 in the NMR structures, respectively. The backbone atoms of the four residues which chelate Zn<sup>2+</sup> are rendered in ball-and-stick while those which are at the dimerization interface are rendered in bonds. The Zn<sup>2+</sup> cation is shown as a ball in all diagrams. Dashed lines show the ion coordination.

buries  $>1800 \text{ \AA}^2$  of molecular surface and consists of mainly polar residues (Lys14, Gln18, Gln44, Lys160, Gln168, Lys186 & Lys188). It is tempting to regard this tetramer as physiologically relevant. However, the authors note that gel filtration experiments at physiological salt concentrations demonstrate that 1-212 is exclusively dimeric [43], raising the question of whether this tetramer is functionally relevant.

### 3. METAL BINDING TO THE INTEGRASE ACTIVE SITE

Integrase requires  $\text{Mn}^{2+}$  or  $\text{Mg}^{2+}$  for catalytic activity [56]. However, there has been some ambiguity about whether there are one or two metal ion binding sites in the enzyme. Divalent metal ions have been observed in several of the crystal structures of the core domains of HIV-1 and ASV integrase [10,12,14,16,25,46].  $\text{Mg}^{2+}$  binds to a single site in HIV-1 integrase, forming a single bond each to a carboxylate oxygen of Asp64 and Asp116. ASV integrase has been shown to bind  $\text{Mn}^{2+}$  and  $\text{Mg}^{2+}$ , and  $\text{Ca}^{2+}$  similarly at a single site (Fig 5). However, both  $\text{Zn}^{2+}$  (which only supports 3' processing) and  $\text{Cd}^{2+}$  (which supports neither 3' processing nor strand transfer) have been observed to bind at two sites, the second site involving Asp64 and Glu157 [16].  $\text{Mn}^{2+}$  and  $\text{Mg}^{2+}$ , but not  $\text{Ca}^{2+}$ , support 3' processing and strand transfer.



**Fig. (5).** Cation coordination by active site residues of ASV integrase core domains [14,16,17]. The side chain of E157 has the same conformation in the cation-free and single cation structures, but coordinates an additional cation in the  $\text{Cd}^{2+}$  and  $\text{Zn}^{2+}$  structure.

In the case of homologous structures such as RNaseH, which also require  $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$  for activity, the results are also not clear. For the *E. coli* enzyme it has been shown that only one  $\text{Mg}^{2+}$  binds in cocrystals grown from a high  $\text{Mg}^{2+}$  concentration [39]. This metal binding site is located in a site similar to that in ASV integrase. However, crystallization of HIV-1 RNaseH with even quite low concentration (1mM) of  $\text{Mn}^{2+}$  results in two binding sites being occupied [57,58]. It seems probable, based on the present evidence that two

metal ions will be bound, with the second being more tightly bound in the presence of substrate DNA. From observations on the active site of the endonuclease from the Klenow fragment of DNA polymerase Beese and Steitz proposed a mechanism of action involving two  $\text{Mg}^{2+}$  ions [59], and it is possible that a similar mechanism is also used by integrase.

### 4. INHIBITORS OF HIV INTEGRASE

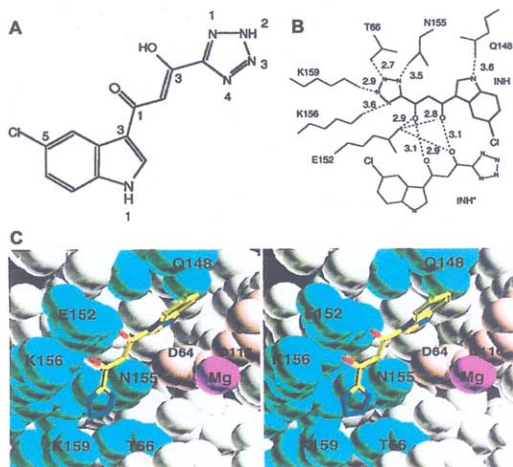
Integrase is an important target for anti- HIV-1 drug design. However, until now the successful development of antiviral agents has primarily involved two enzymes: reverse transcriptase and protease. Many inhibitors of integrase have been discovered, but they have not yet been developed from lead compound to effective therapeutic agents [60-64].

The inhibitors of HIV-1 integrase are quite diverse; they include nucleotides, diketoacids [46,65], polyamides [66], guanine quartets, among others. Of particular interest are the diketoacid inhibitors studied by the Merck group [65,67,68]. These inhibit in particular the strand transfer reaction. They bind only full length integrase, and apparently only when integrase is associated with the ends of the viral DNA. Binding of the diketoacid inhibitor is correlated with inhibition of strand transfer.

#### HIV-1 Core Plus 5-CITEP Structure

There is at present only one crystal structure of the core domain of HIV-1 integrase bound to an inhibitor located centrally within the active site [46]. The inhibitor 1-(5-chloroindole-3-yl)-3-hydroxy-3(2H-tetrazol-5-yl)-propenone (or 5-CITEP, Fig. 6a) has some resemblance to the diketoacid inhibitors, but has a tetrazolium group instead of a carboxylate. This inhibitor, introduced by soaking preformed crystals, binds to one of the three active sites in the asymmetric unit, with only weak uninterpretable electron density at the other two sites. The inhibitor is located between the active site residues Asp64, Asp116 and Glu152 and is bound to the surface by a variety of hydrogen bonds between the tetrazolium ring and neighboring amino acids (Fig. 6bc), some of which (Lys156 and Lys159) have been identified previously [69] as being involved in DNA binding to the core domain. The N1 atom of the indole is also within hydrogen bonding distance of Gln148, which has also been shown to bind the 5' adenine on the strand complementary to the processed strand [70]. The  $\text{Mg}^{2+}$  ion remains bound to the two aspartic acids in the presence of the inhibitor. The active site at which the inhibitor binds is close to a crystallographic two-fold axis and the inhibitor interacts with an identical inhibitor across this two-fold axis. It is possible that this could affect the crystallographically observed mode of binding to the active site. A computational docking analysis [71] supports the binding at the two sites related by the two-fold axis, but the binding to an isolated domain is predicted to be different from that observed in the X-ray analysis.

It is necessary to inject a word of caution here regarding the uniqueness of the integrase as a target for these diketo compounds. Melek *et.al.* 2002 [72] have examined the effect of these inhibitors on the RAG1/2 recombinase associated with VDJ recombination, an enzyme system that carries out



**Fig. (6).** (A). The inhibitor 5-CITEP. (B) The contacts between the inhibitors related by the crystallographic two-fold axis and the side chains of amino acid residues at the active sites. (C) Stereo image of the inhibitor in one of the monomer. Contacting side chains are shown in cyan, with magnesium in purple and asp64 and asp116 in pink. Figure from reference [46].

reactions that are very similar to those of HIV-1 integrase. They observed that both the diketoacid, L-708.906 [67], and 5-CITEP were capable of inhibiting RAG1/2 reactions, although by about a factor of ten less than their effects on HIV-1 integrase. It will therefore be necessary in the future, when looking for inhibitors of integrase with therapeutic potential, to assay their effect on the RAG1/2 recombinase in order to avoid potentially deleterious side effects. However, it should be noted that these are two lead compounds and that the development of more specific inhibitors designed specifically for integrase will probably result in a much greater difference in affinity to related enzymes such as the RAG1/2 recombinase.

#### Structural Studies on Two Other Inhibitors of Integrase

A crystal structure has been reported for HIV-1 integrase complexed with the inhibitor dihydroxyphenyltetraphenylarsonium [45]. Tetraphenylarsonium, a much poorer inhibitor, binds to the same site. The binding site is located distant from the active site at the interface between the two monomers of the core domain dimer. There is also contact with a side chain from a third crystallographically related monomer subunit, raising the possibility that binding at this site might be due to crystal packing factors. Binding of the inhibitor to these trigonal crystals which were grown in cacodylate buffer resulted in a displacement of the covalently bound arsenic atom linked to Cys65, causing it to swing out of the hydrophobic pocket by 3Å.

The second study has the inhibitor 4-acetylamino-5-hydroxynaphthalene-2,7-disulfonic acid bound to ASV integrase [73]. The inhibitor, originally identified in a screen for inhibitors of HIV-1 integrase, also inhibited ASV

integrase. This inhibitor binds to the core domain of ASV integrase, not too distant from the active site, but on the other side of the flexible loop. The inhibitor interacts with a second integrase molecule across a two-fold axis, again raising the possibility that crystal packing plays a role in the binding. Comparison of the amino acid residues making contact with the inhibitor with those in HIV-1 integrase indicates that comparable binding of the inhibitor to the HIV-1 enzyme is possible.

#### Oligonucleotide Inhibitors

Inhibition of both 3' processing and strand transfer occur with either unmodified or modified oligonucleotides, reviewed in reference [63]. Although nonviral oligonucleotides can inhibit 3' processing, DNA containing the viral U5 sequence can inhibit 3' processing to a stronger extent. For example, the  $IC_{50}$  for the 3' processing of a 21mer U5 DNA is 4nM whereas that of deoxy(T:A)<sub>21</sub> is only 320 nM, or 80 times weaker [74]. DNA containing a nonhydrolyzable group at the terminal AG base step also inhibits 3' processing with nanomolar  $IC_{50}$  [75,76] whereas single-stranded DNA of the same sequence has a ten fold weaker inhibition [74]. Shorter sequences, especially mono- and dinucleotides inhibit activity at the micromolar to sub-millimolar range [61,74,77].

#### Polyamide Inhibitors

Polyamides or lexitropsins consisting of contiguous pyrrole rings are compounds which bind the minor groove of B-DNA at AT-rich sequences [78]. Derivatives in which the pyrrole ring has been replaced with one containing an imidazole can bind to GC-containing sequences [78,79]. In addition polyamides containing other substitutions and which are conformationally restrained by covalent crosslinking to bind the DNA minor groove as dimers, can in principal be designed to read any DNA sequence [80-85]. However, the degree of specificity required for a drug to bind only viral DNA and not the 10<sup>9</sup> bases of the human genome (a one in a billion binding is equivalent to recognizing a stretch of 15 bases) is currently not obtainable, in part because the readout of the DNA bases by these drugs get out of register with the DNA helix after about seven base pairs [86,87].

HIV-1 viral DNA has a seven nucleotide stretch of AT-rich sequence ~10 bases away from each LTR end that presents a target for AT-reading polyamides. The  $IC_{50}$  for the 3' processing reaction of distamycin A (which has three pyrrole rings) and polyamides containing four thiazole rings are 3.5µM and 0.3µM, respectively [66]. Bis-distamycin derivatives inhibit 3' processing, strand transfer and disintegration to a stronger extent, and inhibit HIV-1 replication at ~0.1mM concentration [88]. These polyamides, in particular have been shown to inhibit reverse transcriptase *in vitro* [89].

#### G-quartets

Oligonucleotides forming G quartets have been shown to inhibit the integrase reactions *in vitro* with  $IC_{50}$  values approaching the nanomolar range [90,91]. Some of these G quartet sequences also interfere with virus cell binding

and viral entry, and this has been interpreted as binding to GP120 [92,93].

## 5. INTEGRASE – DNA INTERACTIONS

Current experimental evidence from photocrosslinking and complementation studies suggest that all three protein domains interact with DNA: the N-terminal domain is in close proximity to target DNA 5' to the site of integration [94], the core domain interacts with terminal bases of viral LTR ends [56,69,70,94-100] and target DNA flanking the joined site of integration [101], and the C-terminal domain interacts with bases distal to the terminal bases of the LTRs [69,70,94]. Furthermore, complementation studies of domain fragments of HIV-1 and Visna virus [38] integrases, which share about 31% sequence identity, suggests the N-terminal domain is not involved in determining substrate specificity for 3' processing and strand transfer [8]. In particular, the first 26 residues of RSV and HIV-1 integrase, which includes the first two histidines in the His<sub>2</sub>Cys<sub>2</sub> motif, are not required for DNA binding [5,100].

Functional activity in HIV-1 integrase requires a DNA ending with the sequence CAGT. Adding nucleotides to the 3' end of this sequence severely reduces 3' processing [102]. Also, substitution of nucleotide analogs that strengthen the hydrogen bonding between the plus and the minus strand decreases 3' processing activity, while those which weaken or disrupt base pairing in the vicinity of the CA dinucleotide increase activity [100,103-105]. Complementation and crosslinking data also suggest the N-terminal domain acts in trans to the core domain containing a functional active site [33,38,95,106], while the C-terminal domain can be in cis [38] or trans [28,38,95,97] to the core domain. In addition, viral DNA is bound in trans to an active core domain.

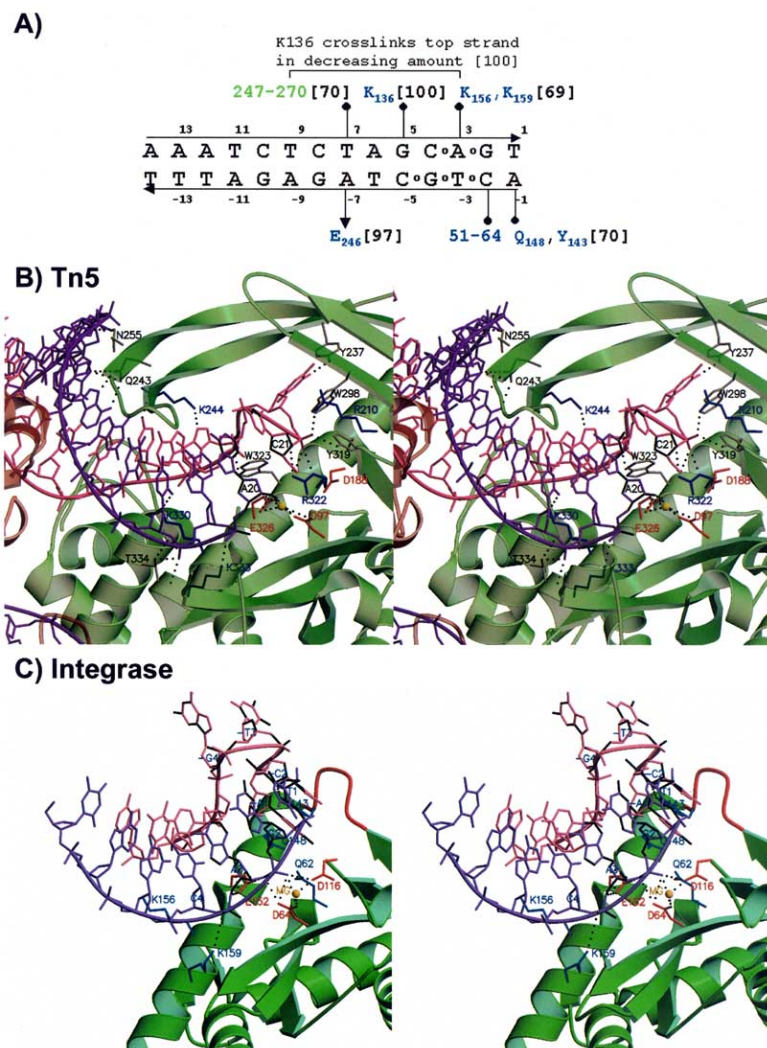
Although the 3' processing and strand transfer activities of retroviral integrase are sequence-specific for the proximal part of the viral LTR end [69,96,99,107], the binding of integrase to DNA is not [100,108,109]. Consistent with this are: mutant LTR substrates which are cleaved inefficiently can still be crosslinked to integrase [99] and inhibit processing of wild type substrate [96]; and integrase's protease cleavage pattern is the same when it is bound to either U5 or non-U5 substrate [110]. Although integrase can bind both single-stranded and double-stranded DNA, the particular interactions appear to be different, as single-stranded DNA, but not double-stranded DNA, when bound to integrase is susceptible to phosphodiesterase cleavage [109]. Binding of integrase to U5 LTR DNA is also cooperative, weaker at higher salt concentration, better with Mn<sup>2+</sup> than with Mg<sup>2+</sup>, and stoichiometrically requires at least four integrase monomers per DNA duplex [109]. For a 21bp DNA, this integrase/DNA/cation complex is also more stable and more resistant to phosphodiesterase cleavage and EDTA inhibition in the presence of Mn<sup>2+</sup> than in the presence of Mg<sup>2+</sup> [100,109]. However, this metal preference is less for a 35mer, and divalent cation is not required for longer sequences [100]. The metal preference is most likely caused by the greater stability of the integrase/DNA/cation complex with Mn<sup>2+</sup> because the rate of cleavage is similar for both cations [109].

Fig. (7a) summarizes the experimental data on protein and DNA residues which have been implicated by crosslinking or mutagenesis to be involved in integrase – DNA interactions. These data, along with the observation that mutation of bases 3(+) – 6(+) decrease 3' processing and abolish strand transfer activities while mutation of bases 7(+) – 13(+) has no effect on either activity [96,99], suggest that residues within the catalytic core are responsible for sequence-specific recognition of the last six base pairs of viral LTR ends while residues within the C-terminal domain are responsible for nonspecific binding of more distal bases. In particular, interactions between the distal nonconserved bases and the C-terminal domain may be responsible for correctly positioning and orientating the viral DNA so that the conserved bases (1 to 6) can be recognized by specific residues of the core domain and catalysis can occur. Indeed, integrase without a C-terminal domain does not have 3' processing and strand transfer activities, presumably because it can not correctly position and orient the viral LTR ends at the active site. However, when the viral DNA ends have been correctly prepositioned, such as with a disintegration substrate, catalysis can occur without the C-terminal and N-terminal domains.

The crystal structure of Tn5 transposase bound to a specific recognition sequence is shown in Fig. (7b) [111]. Although transposition with Tn5 transposase involves a covalent hairpin intermediate instead of the two step pathway of retroviral integration, these two processes are mechanistically similar enough so that the Tn5 transposase – DNA structure can be used as a starting point for modeling retroviral integrase – DNA interactions. The DNA substrate used in the Tn5 study has a blunt end but in the structure the last two bases of each strand are unpaired. The last two bases of the strand with the free 3' hydroxyl group (magenta) have the normal B-DNA conformation while the last two bases of the complementary strand (red) are looped out. In this conformation the phosphate group of the looped out strand (which is missing in the substrate used in the study) is in position for nucleophilic attack by the free 3' hydroxyl; the two bases involved are A20 and C21 in Fig. (7b). The catalytic residues Asp97, Asp188 and Glu326 of the DDE motif as well as the Mn<sup>2+</sup> cation are in position for this step and the DNA is held in place by protein residues Arg210, Tyr237, Gln243, Lys244, Asn255, Trp298, Tyr319, Arg322, Trp323, Lys330, Lys333, and Thr334.

The core domain of HIV-1 integrase can be superimposed onto corresponding residues of Tn5 transposase, and this is shown in Fig. (7c). Residues Lys156 and Lys159, and residues Tyr143 and Gln148, which have been shown to crosslink base A(+3) and A(-1), respectively [69,70], are used to position the viral U5 DNA in a similar orientation at the active site. Lys156 and Lys159 of HIV-1 integrase are equivalent to residues Lys330 and Lys333, respectively of Tn5 transposase. The phosphate group of A(+3), which is one base upstream of the cleaved phosphodiester bond, is placed in the general vicinity of the phosphate ion observed in the N-terminus plus core domain structure [24]. The four terminal bases of the complementary strand (bases -1 to -4), are looped out in a manner similar to that of Tn5 transposase, and the A(-1) base is modelled as being sandwiched in





**Fig. (7).** (A) Protein residues or peptide fragments which crosslink (blue) or interact (green; suggested by effect on activity) with linear HIV-1 U5 DNA. Sugar phosphate backbone where ethylation of HIV-1 U3 or U5 LTR substrate interferes with strand transfer activity are represented by circles [98]. Numbers in brackets gives reference for original data. (B) Stereo image of Tn5 transposase with DNA (coordinate 1F3I) [111], showing closeup of the active site. Backbone atoms of active side residues are in red while other residues which interact with DNA are in blue or grey. Each DNA strand is shown in different color. A small gold ball represents Mn<sup>2+</sup>. Hydrogen bonds are represented by dashed lines. (C) Stereo image of a possible model of HIV-1 integrase with DNA, generated using data in (A) and mirroring the Tn5 transposase – DNA structure. The protein flexible loop is shown in red.

between residues Tyr143 and Gln148. This interaction reflects both crosslinking data and mutagenesis data [70]. Removal of the 5' AC overhang and the Q148L mutation of integrase both destabilize integrase – DNA complexes [56,104]. Deletion of the 5' AC overhang results in inhibition of strand transfer activity by polylysine [56].

Although this end of the viral LTR can be reasonably modelled at the active site, how the other end interacts with integrase is unclear without a structure of full length integrase. As indicated earlier, the C-terminus acts in trans to the catalytic core but how this is structurally achieved is unclear. Similarly, it is not clear how target DNA is arranged in the integrase – DNA complex, although chimeras of HIV-

1 and FIV integrases suggest the core domain is involved in target site selection [8,112].

## CONCLUSION

As we have attempted to show in this review, there is now a considerable structural database for HIV-1 integrase. Structures exist for each of the three domains of the molecule, and for two two-domain constructs, however, more information is needed. We do not have a structure for the whole molecule either with or without DNA substrate, and a detailed structural model for the catalytic mechanism of integration is still not available. The three structures of inhibitor bound to HIV-1 integrase are insufficient to provide

a satisfactory database for lead compound development into therapeutic drugs. Much progress has been made but there is still much to be learned.

## ACKNOWLEDGEMENTS

We thank the National Institutes of Health for an IATAP grant which funded this work. We thank also Dr. Fred Dyda, Dr. Alison Hickman and Dr. Robert Craigie for helpful comments on this manuscript. Figures 2 – 5 and 7 were drawn with Molscript [113], ImageMagick ©, and Raster3d [114].

## REFERENCES

- [1] Brown, P. O. In *Retroviruses*; Coffin, J. M.; Hughes, S. H.; Varmus, H. E.; Eds.; Cold Spring Harbor Laboratory, **1997**; pp 161-203.
- [2] Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *Journal of Biological Chemistry* **2001**, *276*, 23213-23216.
- [3] Asante-Appiah, E.; Skalka, A. M. Molecular mechanisms in retrovirus DNA integration. *Antiviral Research* **1997**, *36*, 139-156.
- [4] Hindmarsh, P.; Leis, J. Retroviral DNA integration. *Microbiology and Molecular Biology Reviews* **1999**, *63*, 836-843.
- [5] Mumm, S. R.; Grandgenett, D. P. Defining Nucleic Acid-Binding Properties of Avian Retrovirus Integrase by Deletion Analysis. *Journal of Virology* **1991**, *65*, 1160-1167.
- [6] Engelman, A.; Craigie, R. Identification of Conserved Amino-Acid-Residues Critical for Human-Immunodeficiency-Virus Type-1 Integrase Function-Invitro. *Journal of Virology* **1992**, *66*, 6361-6369.
- [7] Vangent, D. C.; Groeneger, A. M. O.; Plasterk, R. H. A. Identification of Amino-Acids in Hiv-2 Integrase Involved in Site-Specific Hydrolysis and Alcoholysis of Viral-DNA Termini. *Nucleic Acids Research* **1993**, *21*, 3373-3377.
- [8] Katzman, M.; Sudol, M. Mapping Domains of Retroviral Integrase Responsible for Viral- DNA Specificity and Target Site Selection by Analysis of Chimeras between Human-Immunodeficiency-Virus Type-1 and Visna Virus Integrases. *Journal of Virology* **1995**, *69*, 5687-5696.
- [9] Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. Crystal-Structure of the Catalytic Domain of Hiv-1 Integrase - Similarity to Other Polynucleotidyl Transferases. *Science* **1994**, *266*, 1981-1986.
- [10] Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 9150-9154.
- [11] Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* **1999**, *38*, 8892-8898.
- [12] Maignan, S.; Guilloreau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: High level of similarity of the active site with other viral integrases. *Journal of Molecular Biology* **1998**, *282*, 359-368.
- [13] Bujacz, G.; Jaskoski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. High-Resolution Structure of the Catalytic Domain of Avian- Sarcoma Virus Integrase. *Journal of Molecular Biology* **1995**, *253*, 333-346.
- [14] Bujacz, G.; Jaskolski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. The catalytic domain of avian sarcoma virus integrase: Conformation of the active-site residues in the presence of divalent cations. *Structure* **1996**, *4*, 89-96.
- [15] Bujacz, G.; Alexandratos, J.; Zhou-Liu, Q.; Clement-Mella, C.; Wlodawer, A. The catalytic domain of human immunodeficiency virus integrase: Ordered active site in the F185H mutant. *Febs Letters* **1996**, *398*, 175-178.
- [16] Bujacz, G.; Alexandratos, J.; Wlodawer, A. Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *Journal of Biological Chemistry* **1997**, *272*, 18161-18168.
- [17] Lubkowski, J.; Yang, F.; Alexandratos, J.; Merkel, G.; Katz, R. A.; Gravuer, K.; Skalka, A. M.; Wlodawer, A. Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase. *Journal of Biological Chemistry* **1998**, *273*, 32685-32689.
- [18] Lubkowski, J.; Dauter, Z.; Yang, F.; Alexandratos, J.; Merkel, G.; Skalka, A. M.; Wlodawer, A. Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant (vol 38, pg 13512, 1999). *Biochemistry* **1999**, *38*, 15060-15060.
- [19] Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution Structure of the DNA-Binding Domain of Hiv-1 Integrase. *Biochemistry* **1995**, *34*, 9826-9833.
- [20] Eijkelenboom, A.; Sprangers, R.; Hard, K.; Lutzke, R. A. P.; Plasterk, R. H. A.; Boelens, R.; Kaptein, R. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins-Structure Function and Genetics* **1999**, *36*, 556-564.
- [21] Cai, M. L.; Zheng, R. L.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Structural Biology* **1997**, *4*, 567-577.
- [22] Cai, M. L.; Huang, Y.; Caffrey, M.; Zheng, R. L.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the His12 -> Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium. *Protein Science* **1998**, *7*, 2669-2674.
- [23] Eijkelenboom, A.; vandenEnt, F. M. I.; Vos, A.; Doreleijers, J. F.; Hard, K.; Tullius, T. D.; Plasterk, R. H. A.; Kaptein, R.; Boelens, R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Current Biology* **1997**, *7*, 739-746.
- [24] Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO Journal* **2001**, *20*, 7333-7343.
- [25] Chen, J. C. H.; Krucinski, J.; Miercke, L. J. W.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 8233-8238.
- [26] Chen, Z. G.; Yan, Y. W.; Munshi, S.; Li, Y.; Zuyag-Murphy, J.; Xu, B.; Witmer, M.; Felock, P.; Wolfe, A.; Sardana, V.; Emini, E. A.; Hazuda, D.; Kuo, L. C. X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293) - An initial glance of the viral DNA binding platform. *Journal of Molecular Biology* **2000**, *296*, 521-533.
- [27] Yang, Z. N.; Mueser, T. C.; Bushman, F. D.; Hyde, C. C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *Journal of Molecular Biology* **2000**, *296*, 535-548.
- [28] Vangent, D. C.; Vink, C.; Groeneger, A.; Plasterk, R. H. A. Complementation between Hiv Integrase Proteins Mutated in Different Domains. *EMBO Journal* **1993**, *12*, 3261-3267.
- [29] Kulkosky, J.; Jones, K. S.; Katz, R. A.; Mack, J. P. G.; Skalka, A. M. Residues Critical for Retroviral Integrative Recombination in a Region That Is Highly Conserved among Retroviral Retrotransposon Integrases and Bacterial Insertion-Sequence Transposases. *Molecular and Cellular Biology* **1992**, *12*, 2331-2338.
- [30] Drelich, M.; Wilhelm, R.; Mous, J. Identification of Amino-Acid-Residues Critical for Endonuclease and Integration Activities of Hiv-1 in Protein Invitro. *Virology* **1992**, *188*, 459-468.
- [31] Bushman, F. D.; Engelman, A.; Palmer, I.; Wingfield, P.; Craigie, R. Domains of the Integrase Protein of Human-Immunodeficiency-Virus Type-1 Responsible for Polynucleotidyl Transfer and Zinc-Binding. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 3428-3432.
- [32] Mazumder, A.; Engelman, A.; Craigie, R.; Fesen, M.; Pommier, Y. Intermolecular Disintegration and Intramolecular Strand Transfer Activities of Wild-Type and Mutant Hiv-1 Integrase. *Nucleic Acids Research* **1994**, *22*, 1037-1043.
- [33] Jonsson, C. B.; Donzella, G. A.; Gaucan, E.; Smith, C. M.; Roth, M. J. Functional domains of moloney murine leukemia virus

- integrane defined by mutation and complementation analysis. *Journal of Virology* **1996**, *70*, 4585-4597.
- [34] Shibagaki, Y.; Holmes, M. L.; Appa, R. S.; Chow, S. A. Characterization of feline immunodeficiency virus integrase and analysis of functional domains. *Virology* **1997**, *230*, 1-10.
- [35] Vink, C.; Groeneger, A.; Plasterk, R. H. A. Identification of the Catalytic and DNA-Binding Region of the Human-Immunodeficiency-Virus Type-1 Integrase Protein. *Nucleic Acids Research* **1993**, *21*, 1419-1425.
- [36] Schauer, M.; Billich, A. The N-Terminal Region of Hiv-1 Integrase Is Required for Integration Activity, but Not for DNA-Binding. *Biochemical and Biophysical Research Communications* **1992**, *185*, 874-880.
- [37] Kulkosky, J.; Katz, R. A.; Merkel, G.; Skalka, A. M. Activities and Substrate-Specificity of the Evolutionarily Conserved Central Domain of Retroviral Integrase. *Virology* **1995**, *206*, 448-456.
- [38] Engelman, A.; Bushman, F. D.; Craigie, R. Identification of Discrete Functional Domains of Hiv-1 Integrase and Their Organization within an Active Multimeric Complex. *EMBO Journal* **1993**, *12*, 3269-3275.
- [39] Katayanagi, K.; Okumura, M.; Morikawa, K. Crystal-Structure of Escherichia-Coli Rnase Hi in Complex with Mg<sup>2+</sup> at 2.8 Angstrom Resolution - Proof for a Single Mg<sup>2+</sup>-Binding Site. *Proteins-Structure Function and Genetics* **1993**, *17*, 337-346.
- [40] Rice, P.; Mizuuchi, K. Structure of the Bacteriophage-Mu Transposase Core - a Common Structural Motif for DNA Transposition and Retroviral Integration. *Cell* **1995**, *82*, 209-220.
- [41] Ariyoshi, M.; Vassilyev, D. G.; Iwasaki, H.; Nakamura, H.; Shinagawa, H.; Morikawa, K. Atomic-Structure of the RuvC Resolvase - a Holliday Junction-Specific Endonuclease from Escherichia-Coli. *Cell* **1994**, *78*, 1063-1072.
- [42] Jenkins, T. M.; Hickman, A. B.; Dyda, F.; Ghirlando, R.; Davies, D. R.; Craigie, R. Catalytic Domain of Human-Immunodeficiency-Virus Type-1 Integrase - Identification of a Soluble Mutant by Systematic Replacement of Hydrophobic Residues. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 6057-6061.
- [43] Jenkins, T. M.; Engelman, A.; Ghirlando, R.; Craigie, R. A soluble active mutant of HIV-1 integrase - Involvement of both the core and carboxyl-terminal domains in multimerization. *Journal of Biological Chemistry* **1996**, *271*, 7712-7718.
- [44] Engelman, A.; Liu, Y.; Chen, H. M.; Farzan, M.; Dyda, F. Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. *Journal of Virology* **1997**, *71*, 3507-3514.
- [45] Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallographica Section D-Biological Crystallography* **2001**, *57*, 536-544.
- [46] Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 13040-13043.
- [47] Lubkowski, J.; Dauter, Z.; Yang, F.; Alexandratos, J.; Merkel, G.; Skalka, A. M.; Wlodawer, A. Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry* **1999**, *38*, 13512-13522.
- [48] Bushman, F. D.; Wang, B. B. Rous-Sarcoma Virus Integrase Protein - Mapping Functions for Catalysis and Substrate-Binding. *Journal of Virology* **1994**, *68*, 2215-2223.
- [49] Pahl, A.; Flugel, R. M. Characterization of the Human Spuma Retrovirus Integrase by Site-Directed Mutagenesis, by Complementation Analysis, and by Swapping the Zinc-Finger Domain of Hiv-1. *Journal of Biological Chemistry* **1995**, *270*, 2957-2966.
- [50] van Gent, D. C.; Vink, C.; Groeneger, A.; Plasterk, R. H. A. Complementation between Hiv Integrase Proteins Mutated in Different Domains. *EMBO Journal* **1993**, *12*, 3261-3267.
- [51] Zheng, R. L.; Jenkins, T. M.; Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 13659-13664.
- [52] Katz, R. A.; Merkel, G.; Skalka, A. M. Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: *In vitro* activities and incorporation of a fusion protein into viral particles. *Virology* **1996**, *217*, 178-190.
- [53] Khan, E.; Mack, J. P. G.; Katz, R. A.; Kulkosky, J.; Skalka, A. M. Retroviral Integrase Domains - DNA-Binding and the Recognition of Ltr Sequences. *Nucleic Acids Research* **1991**, *19*, 851-860.
- [54] Lee, S. P.; Xiao, J. M.; Knutson, J. R.; Lewis, M. S.; Han, M. K. Zn<sup>2+</sup> promotes the self-association of human immunodeficiency virus type-1 integrase *in vitro*. *Biochemistry* **1997**, *36*, 173-180.
- [55] Hickman, A. B.; Palmer, I.; Engelman, A.; Craigie, R.; Wingfield, P. Biophysical and Enzymatic-Properties of the Catalytic Domain of Hiv-1 Integrase. *Journal of Biological Chemistry* **1994**, *269*, 29279-29287.
- [56] Ellison, V.; Brown, P. O. A Stable Complex between Integrase and Viral-DNA Ends Mediates Human-Immunodeficiency-Virus Integration *in-Vitro*. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 7316-7320.
- [57] Goedken, E. R.; Keck, J. L.; Berger, J. M.; Marqusee, S. Divalent metal cofactor binding in the kinetic folding trajectory of Escherichia coli ribonuclease HI. *Protein Science* **2000**, *9*, 1914-1921.
- [58] Davies, J. F.; Hostomska, Z.; Hostomsky, Z.; Jordan, S. R.; Matthews, D. A. Crystal-Structure of the Ribonuclease-H Domain of Hiv-1 Reverse-Transcriptase. *Science* **1991**, *252*, 88-95.
- [59] Beese, L. S.; Steitz, T. A. Structural Basis for the 3'-5' Exonuclease Activity of Escherichia-Coli DNA-Polymerase-I - a 2 Metal-Ion Mechanism. *EMBO Journal* **1991**, *10*, 25-33.
- [60] Neamati, N.; Turpin, J. A.; Winslow, H. E.; Christensen, J. L.; Williamson, K.; Orr, A.; Rice, W. G.; Pommier, Y.; Garofalo, A.; Brizzi, A.; Campiani, G.; Fiorini, I.; Nacci, V. Thiazolothiazepine inhibitors of HIV-1 integrase. *Journal of Medicinal Chemistry* **1999**, *42*, 3334-3341.
- [61] Pommier, Y.; Marchand, C.; Neamati, N. Retroviral integrase inhibitors year 2000: update and perspectives. *Antiviral Research* **2000**, *47*, 139-148.
- [62] Tarrago-Litvak, L.; Andreola, M. L.; Fournier, M.; Nevinsky, G. A.; Parissi, V.; de Soultrait, V. R.; Litvak, S. Inhibitors of HIV-1 reverse transcriptase and integrase: Classical and emerging therapeutic approaches. *Current Pharmaceutical Design* **2002**, *8*, 595-614.
- [63] Nair, V. HIV integrase as a target for antiviral chemotherapy. *Reviews in Medical Virology* **2002**, *12*, 179-193.
- [64] Clercq, E. D. New developments in anti-HIV chemotherapy. *Biochimica et Biophysica Acta* **2002**, *1587*, 258-275.
- [65] Grobler, J. A.; Stillmock, K.; Hu, B. H.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, D.; Hazuda, D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 6661-6666.
- [66] Ryabinin, V. A.; Sinyakov, A. N.; de Soultrait, V. R.; Caumont, A.; Parissi, V.; Zakharova, O. D.; Vasyutina, E. L.; Yurchenko, E.; Bayandin, R.; Litvak, S.; Tarrago-Litvak, L.; Nevinsky, G. A. Inhibition of HIV-1 integrase-catalysed reaction by new DNA minor groove ligands: the oligo-1,3-thiazolecarboxamide derivatives. *European Journal of Medicinal Chemistry* **2000**, *35*, 989-1000.
- [67] Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **2000**, *287*, 646-650.
- [68] Espeseth, A. S.; Felock, P.; Wolfe, A.; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 11244-11249.
- [69] Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO Journal* **1997**, *16*, 6849-6859.

- [70] Esposito, D.; Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *Embo Journal* **1998**, *17*, 5832-5843.
- [71] Sotriffer, C. A.; Ni, H. H.; McCammon, J. A. HIV-1 integrase inhibitor interactions at the active site: Prediction of binding modes unaffected by crystal packing. *Journal of the American Chemical Society* **2000**, *122*, 6136-6137.
- [72] Melek, M.; Jones, J. M.; O'Dea, M. H.; Pais, G.; Burke, T. R.; Pommier, Y.; Neamati, N.; Gellert, M. Effect of HIV integrase inhibitors on the RAG1/2 recombinase. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 134-137.
- [73] Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T. R.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A. M. Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 4831-4836.
- [74] Caumont, A.; Jamieson, G.; de Soultrait, V. R.; Parissi, V.; Fournier, M.; Zakharova, O. D.; Bayandin, R.; Litvak, S.; Tarrago-Litvak, L.; Nevinsky, G. A. High affinity interaction of HIV-1 integrase with specific and non-specific single-stranded short oligonucleotides. *FEBS Letters* **1999**, *455*, 154-158.
- [75] Tramontano, E.; La Colla, P.; Cheng, Y. C. Biochemical characterization of the HIV-1 integrase 3' processing activity and its inhibition by phosphorothioate oligonucleotides. *Biochemistry* **1998**, *37*, 7237-7243.
- [76] Snasel, J.; Rejman, D.; Liboska, R.; Tocik, Z.; Ruml, T.; Rosenberg, I.; Pichova, I. Inhibition of HIV-1 integrase by modified oligonucleotides derived from U5' LTR. *European Journal of Biochemistry* **2001**, *268*, 980-986.
- [77] Taktakishvili, M.; Neamati, N.; Pommier, Y.; Nair, V. Recognition and inhibition of HIV integrase by a novel dinucleotide. *Bioorganic & Medicinal Chemistry Letters* **2000**, *10*, 249-251.
- [78] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The Molecular-Origin of DNA Drug Specificity in Netropsin and Distamycin. *Proceedings of the National Academy of Sciences of the United States of America* **1985**, *82*, 1376-1380.
- [79] Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. Molecular Recognition between Oligopeptides and Nucleic-Acids - Novel Imidazole-Containing Oligopeptides Related to Netropsin That Exhibit Altered DNA-Sequence Specificity. *Biochemistry* **1986**, *25*, 7408-7416.
- [80] Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. Binding of 2 Distamycin-a Molecules in the Minor-Groove of an Alternating B-DNA Duplex. *Nature Structural Biology* **1994**, *1*, 169-175.
- [81] Chen, Y. H.; Lown, J. W. A New DNA Minor-Groove Binding Motif - Cross-Linked Lexitropsins. *Journal of the American Chemical Society* **1994**, *116*, 6995-7005.
- [82] Mrksich, M.; Dervan, P. B. Antiparallel Side-by-Side Heterodimer for Sequence-Specific Recognition in the Minor Groove of DNA by a Distamycin 1- Methylimidazole-2-Carboxamide-Netropsin Pair. *Journal of the American Chemical Society* **1993**, *115*, 2572-2576.
- [83] Mrksich, M.; Dervan, P. B. Enhanced Sequence-Specific Recognition in the Minor-Groove of DNA by Covalent Peptide Dimers - Bis(Pyridine-2- Carboxamidonetropsin)(Ch2)3-6. *Journal of the American Chemical Society* **1993**, *115*, 9892-9899.
- [84] Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A structural basis for recognition of A center dot T and T center dot A base pairs in the minor groove of B-DNA. *Science* **1998**, *282*, 111-115.
- [85] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **1998**, *391*, 468-471.
- [86] Goodsell, D.; Dickerson, R. E. Isohelical Analysis of DNA Groove-Binding Drugs. *Journal of Medicinal Chemistry* **1986**, *29*, 727-733.
- [87] Kelly, J. J.; Baird, E. E.; Dervan, P. B. Binding site size limit of the 2:1 pyrrole-imidazole polyamide- DNA motif. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 6981-6985.
- [88] Neamati, N.; Mazumder, A.; Sunder, S.; Owen, J. M.; Tandon, M.; Lown, J. W.; Pommier, Y. Highly potent synthetic polyamides, bisdistamycins, and lexitropsins as inhibitors of human immunodeficiency virus type 1 integrase. *Molecular Pharmacology* **1998**, *54*, 280-290.
- [89] Filipowsky, M. E.; Kopka, M. L.; BrazilZison, M.; Lown, J. W.; Dickerson, R. E. Linked lexitropsins and the *in vitro* inhibition of HIV-1 reverse transcriptase RNA-directed DNA polymerization: A novel induced-fit of 3,5 m-pyridyl bisdistamycin to enzyme-associated template-primer. *Biochemistry* **1996**, *35*, 15397-15410.
- [90] Mazumder, A.; Neamati, N.; Ojwang, J. O.; Sunder, S.; Rando, R. F.; Pommier, Y. Inhibition of the human immunodeficiency virus type I integrase by guanosine quartet structures. *Biochemistry* **1996**, *35*, 13762-13771.
- [91] Jing, N. J.; Marchand, C.; Liu, J.; Mitra, R.; Hogan, M. E.; Pommier, Y. Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides *in vitro*. *Journal of Biological Chemistry* **2000**, *275*, 21460-21467.
- [92] Este, J. A.; Cabrera, C.; Schols, D.; Cherepanov, P.; Gutierrez, A.; Witvrouw, M.; Pannecouque, C.; Debyser, Z.; Rando, R. F.; Clotet, B.; Desmyter, J.; De Clercq, E. Human immunodeficiency virus glycoprotein gp120 as the primary target for the antiviral action of AR177 (Zintevir). *Molecular Pharmacology* **1998**, *53*, 340-345.
- [93] Jing, N.; De Clercq, E.; Rando, R. F.; Pallansch, L.; Lackman-Smith, C.; Lee, S.; Hogan, M. E. Stability-activity relationships of a family of G-tetrad forming oligonucleotides as potent HIV inhibitors - A basis for anti-HIV drug design. *Journal of Biological Chemistry* **2000**, *275*, 3421-3430.
- [94] Heuer, T. S.; Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **1997**, *36*, 10655-10665.
- [95] Heuer, T. S.; Brown, P. O. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* **1998**, *37*, 6667-6678.
- [96] Lafemina, R. L.; Callahan, P. L.; Cordingley, M. G. Substrate-Specificity of Recombinant Human-Immunodeficiency- Virus Integrase Protein. *Journal of Virology* **1991**, *65*, 5624-5630.
- [97] Gao, K.; Butler, S. L.; Bushman, F. Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO Journal* **2001**, *20*, 3565-3576.
- [98] Bushman, F. D.; Craigie, R. Integration of Human-Immunodeficiency-Virus DNA - Adduct Interference Analysis of Required DNA Sites. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 3458-3462.
- [99] Yoshinaga, T.; Kimurahtani, Y.; Fujiwara, T. Detection and Characterization of a Functional Complex of Human-Immunodeficiency-Virus Type-1 Integrase and Its DNA Substrate by Uv Cross-Linking. *Journal of Virology* **1994**, *68*, 5690-5697.
- [100] Mazumder, A.; Neamati, N.; Pilon, A. A.; Sunder, S.; Pommier, Y. Chemical trapping of ternary complexes of human immunodeficiency virus type I integrase, divalent metal, and DNA substrates containing an abasic site - Implications for the role of lysine 136 in DNA binding. *Journal of Biological Chemistry* **1996**, *271*, 27330-27338.
- [101] Gerton, J. L.; Brown, P. O. The core domain of HIV-1 integrase recognizes key features of its DNA substrates. *Journal of Biological Chemistry* **1997**, *272*, 25809-25815.
- [102] Vink, C.; Vangent, D. C.; Elgersma, Y.; Plasterk, R. H. A. Human-Immunodeficiency-Virus Integrase Protein Requires a Subterminal Position of Its Viral-DNA Recognition Sequence for Efficient Cleavage. *Journal of Virology* **1991**, *65*, 4636-4644.
- [103] Scottoline, B. P.; Chow, S.; Ellison, V.; Brown, P. O. Disruption of the terminal base pairs of retroviral DNA during integration. *Genes & Development* **1997**, *11*, 371-382.
- [104] Gerton, J. L.; Ohgi, S.; Olsen, M.; Derisi, J.; Brown, P. O. Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *Journal of Virology* **1998**, *72*, 5046-5055.
- [105] Wang, T.; Balakrishnan, M.; Jonsson, C. B. Major and minor groove contacts in retroviral integrase-LTR interactions. *Biochemistry* **1999**, *38*, 3624-3632.
- [106] Ellison, V.; Gerton, J.; Vincent, K. A.; Brown, P. O. An Essential Interaction between Distinct Domains of Hiv-1 Integrase Mediates Assembly of the Active Multimer. *Journal of Biological Chemistry* **1995**, *270*, 3320-3326.



- [107] Vangent, D. C.; Elgersma, Y.; Bolk, M. W. J.; Vink, C.; Plasterk, R. H. A. DNA-Binding Properties of the Integrase Proteins of Human Immunodeficiency Viruses Type-1 and Type-2. *Nucleic Acids Research* **1991**, *19*, 3821-3827.
- [108] Engelman, A.; Hickman, A. B.; Craigie, R. The Core and Carboxyl-Terminal Domains of the Integrase Protein of Human-Immunodeficiency-Virus Type-1 Each Contribute to Nonspecific DNA-Binding. *Journal of Virology* **1994**, *68*, 5911-5917.
- [109] Pemberton, I. K.; Buckle, M.; Buc, H. The metal ion-induced cooperative binding of HIV-1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II). *Journal of Biological Chemistry* **1996**, *271*, 1498-1506.
- [110] Dirac, A. M. G.; Kjems, J. Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *European Journal of Biochemistry* **2001**, *268*, 743-751.
- [111] Davies, D. R.; Goryshin, I. Y.; Reznikoff, W. S.; Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **2000**, *289*, 77-85.
- [112] Shibagaki, Y.; Chow, S. A. Central core domain of retroviral integrase is responsible for target site selection. *Journal of Biological Chemistry* **1997**, *272*, 8361-8369.
- [113] Kraulis, P. J. Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *Journal of Applied Crystallography* **1991**, *24*, 946-950.
- [114] Merritt, E. A.; Murphy, M. E. P. Raster3d Version-2.0 - a Program for Photorealistic Molecular Graphics. *Acta Crystallographica Section D-Biological Crystallography* **1994**, *50*, 869-873.